Comments on M.C. Kennedy & A. O'Hagan

Kleijnen, J.P.C.

Published in:
Journal of the Royal Statistical Society, Series B

Publication date:
2001

A:\JRSS03.wpd

Comments on M.C. Kennedy & A. O'Hagan's 'Bayesian calibration of computer models', by

Professor Jack P.C. Kleijnen, Tilburg University, Tilburg, the Netherlands (kleijnen@kub.nl)
Tilburg University, Postbox 90153, 5000 LE Tilburg, the Netherlands

I am a *non-Bayesian* analyst of *stochastic discrete-event simulations*. Such simulations represent 'residual variability' (see §2.1.3, §4.6) through *(pseudo)random numbers*. Examples are queueing simulations in logistics, which generate customers' arrival times through a Poisson distribution with a fixed, unknown parameter θ, so waiting times become random outputs.

In such simulations, *calibration* is considered bad practice! For example, observed arrival times should be used to fit an input distribution. Also see the case-study for the Dutch navy in Kleijnen (2000).

I agree that - after estimating the input parameters - uncertainty remains.. But, this uncertainty may be modeled through *Monte Carlo* sampling, using a fitted input distribution; see §2.3. Alternatively, *bootstrapping* may be used; see Kleijnen (2000). Neither approach assumes *normality* for the input or output distributions!

Code uncertainty (§2.1.6) may be caused by programming *bugs*, see verification - not validation - in Kleijnen (2000).

Instead of assuming a *prior* distribution on the linear regression parameters β, we may first estimate β, and then check whether these estimates have signs that agree with experts' prior knowledge. For example, in queueing simulations the estimated main effect of traffic rate θ on waiting time should be positive; else the model violates validation or verification. In a (deterministic) 'global greenhouse' simulation we detected that two computer modules were called in the wrong order.

I think that the *multivariate* character of the code output (§4.1) is relevant: their correlations should be incorporated through generalized least squares. These correlations can be estimated from replicated runs.

Besides *experimental design* for deterministic simulation (§5.1), there are designs (including screening) for stochastic simulation; see Kleijnen (1998).

Observed input data in historical order can validate the total model, so-called *trace-driven* simulation; see Kleijnen et al. (2000).

We should avoid *spurious regressors* (§5.2), in parametric regression too: though they improve the fit, they also deteriorate the predictor. Moreover, besides *prediction* regression serves (parsimonious) *explanation*!

Bayesians and frequentists - and deterministic and stochastic simulationists - should learn from each other: 'East is East, ...' should not apply to the computer simulation area! Fortunately, the authors have succeeded in writing an article that challenges researchers with diverse backgrounds.

References

Kleijnen, J.P.C. (2000), Strategic directions in verification, validation, and accreditation research: a personal view. *Proceedings 2000 Winter Simulation Conference,* pp. 909-916

---(1998), Experimental design for sensitivity analysis, optimization, and validation of simulation models. *Handbook of Simulation*, edited by J. Banks, Wiley, New York, pp.

173-223